[1] Z. Zhong, M. Hirano, K. Shimada, K. Tateishi, S. Takahashi, and Y. Mitsufuji, “An Attention-Based Approach to Hierarchical Multi-Label Music Instrument Classification,” arXiv (Cornell University), Jun. 2023, doi: <https://doi.org/10.1109/icassp49357.2023.10095162>

**What they did:**

The researchers developed an attention-based approach for hierarchical multi-label music instrument classification. They created a tone-based hierarchical dataset from the OpenMIC dataset, which includes 10-second music clips from the FMA dataset annotated for 20 instruments in a multi-label manner. They introduced a 2-level instrument hierarchy based on tonal properties and used this dataset to train deep neural networks (DNNs) for hierarchical classification.

**Data collection:**

The data for the study came from the [OpenMIC Dataset 2018](https://github.com/cosmir/openmic-2018), which provides 10-second music clips from various genres. These clips are annotated for the presence of 20 different musical instruments in a multi-label fashion. The researchers augmented the dataset by introducing a hierarchical structure based on the tonal properties of the instruments.

**Smart algorithms:**

The researchers proposed two novel methods for joint training of DNNs: one using rule-based grouped max-pooling (GMP) and another employing a data-driven attention mechanism called ResAtt. These methods aim to model the connection between fine- and coarse-level tags, allowing for better interpretation of the decision procedure.

The audio data were converted to mono-channel with a 16kHz sampling rate and transformed into 64-bin mel-spectrograms. The frequency range was set between 50Hz and 8kHz, using a short-time Fourier transform with a 32-ms Hann window and a 10-ms hop size.

**Testing & improvement:**

The researchers evaluated their proposed methods against several baselines and existing approaches. They used metrics such as ROC-AUC, PR-AUC, and F1 score for objective evaluation. The study involved training with different random seeds and hyperparameter tuning to optimize performance.

**Why it works well:**

The proposed joint training methods, particularly the ResAtt approach, showed advantages over traditional methods without joint training. The ResAtt method can learn flexible aggregation rules and prevent fine-level errors from propagating to the coarse level. This leads to improved performance and interpretability, making it a promising approach for hierarchical multi-label music instrument classification tasks.

**Future scope / Drawbacks:**

We have to extend current methods to other tasks in music tagging.

‌